

""" src2doc.py – Takes a source file and makes a document file. Default the source is expected to be Python, and the generated document will be L^AT_EX.

Unless other notices are present in any part of this file explicitly claiming copyrights for other people and/or organizations, the contents of this file is fully copyright (C) 1997 Norut IT, all rights reserved.

Permission is granted to make and distribute verbatim copies of this document provided that the copyright notice and this permission notice are preserved on all copies.

Permission is granted to copy and distribute modified versions of this document under the conditions for verbatim copying, provided that the entire resulting derived work is distributed under the terms of a permission notice identical to this one.

If any other present copyright notices interfere with the copyright claims above, these claims may be partially overruled by those notices. """

```

rcsFile = "$RCSfile: src2doc.py,v $"
rcsDate = "$Date: 2001-08-09 12:33:27 $"
rcsRev = "$Revision: 1.2 $"
rcsState = "$State: Exp $"
rcsAuthor = "$Author: aa $"
rcsLog = ""
$Log: src2doc.py,v $
Revision 1.2 2001-08-09 12:33:27 aa
Tupple stuff (changes in Python between 1.5.2 and 2.0.1?)

Revision 1.1 2000/07/25 13:42:14 aa
Moved to notebook (aa)

Revision 0.4 1997/11/29 18:45:56 anders
Added support for raw string format (r|R) in Python 1.5.

Revision 0.3 1997/10/17 22:10:13 anders
Changed the doc string cleaned up the layout for the Src and Doc class.

Revision 0.2 1997/07/10 13:35:45 anders
Made all values in option dictionaries strings.

Revision 0.1 1997/05/26 00:01:37 anders
Initial working version based on src2ltx v0.1.
"""

```

""" We are importing some library modules. We need `regex` because we are using regular expression to describe source tokens. The `string` module gives us functions for string manipulation. We need the `stdio` stuff from `sys` and we use `types` to know the types of some objects. """

```

import regex # Use the new "re" next time
import string
import sys
from types import *

```

""" An exception used when we don't find a matching component in the source. """

```
NoSymbolMatch = "NoSymbolMatch"
```

```
class Src:
```

""" The Src class is the specification of the source. It must contain two lists of two-tuples; `srctok`, which specifies the tokens to be recognized, and `srccomp`, which specifies the language components to be recognized (language components are made from tokens in `srctok`). It also contains an option dictionary.

The `ws` token must be defined in `srctok`. We also have six tokens that are not defined in `srctok`; `bod` (beginning of document), `eod` (end of document), `bof` (beginning of file), `eof` (end of file), `bol` (beginning of line) and `eol` (end of line). These are automatically inserted in the input stream. """

```

""" The option dictionary for Src (with default values). """
option = {
    "startline": "1",          # Usual starts at first line
    "stopline": "-1",         # and stops at last line (-1).
    "tabsize": "8",           # Deafault tabular size is 8.
    "currcomp": ""}

""" Used in the token definitions. """
whitespaces = "[\011\013\014 ]+"
symbolstr = "[-:;\.<>\+ \* / ! % , | & ^ ~ ]"
symbols = "(\\|\\'|\\\"|\\{|\\}|\\#)"
name = "[a-zA-Z_][a-zA-Z0-9_]*"
escape = "\\|\\(|" + symbols + "\\|" + symbolstr + "\\|[a-zA-Z0-9]+\\)"
numbers = "0x[" + string.hexdigits + "]+)"
numbers = numbers + "\\|[0-9]*\\.?[0-9]+L?(e-?[0-9]+\\)?"
keywords = "access\\|and\\|break\\|class\\|continue\\|def\\|del\\|elif"
keywords = keywords + "\\|else\\|except\\|exec\\|finally\\|for\\|from"
keywords = keywords + "\\|global\\|if\\|import\\|in\\|is\\|lambda\\|not"
keywords = keywords + "\\|or\\|pass\\|print\\|raise\\|return\\|try\\|while"

""" Tokens recognized in the source language. """
srctok = [
    ("esc", (escape, [])),
    ("name",
     (name, [
         ("kw",
          (keywords, [
              ("def", ("def", [])),
              ("cls", ("class", [])),
              ("from", ("from", [])),
              ("import", ("import", []))])),
         ("r", ("r\\R", []))])),
    ("num", (numbers, [])),
    ("ws", (whitespaces, [])),
    ("symstr",
     (symbolstr, [
         ("period", ("\\.", [])),
         ("comma", ("\\,", [])),
         ("asterix", ("\\*", []))])),
    ("sym",
     (symbols, [
         ("quote", ("'", [])),
         ("dquote", ("\"", [])),
         ("bquote", ("`", [])),
         ("comchar", ("#", []))]))]

""" Source components recognized. We are using list and not dictionary because the order is important
(we search for components in this order). """
srccomp = [
    ("docstring", [
        ("#", "bol"),
        ("?", "ws"),
        ("%", [
            ("?", "r"),
            ("", "quote"),
            ("", "quote"),
            ("", "quote")]),
        ("*%", [

```

```

        ("^", [
            ("", "quote"),
            ("", "quote"),
            ("", "quote")]]),
    ("%", [
        ("", "quote"),
        ("", "quote"),
        ("", "quote")]),
    ("%", [
        ("?", "ws"),
        ("", "eol")]]),
("docstring", [
    ("#", "bol"),
    ("?", "ws"),
    ("%", [
        ("?", "r"),
        ("", "dquote"),
        ("", "dquote"),
        ("", "dquote")]),
    ("*%", [
        ("^", [
            ("", "dquote"),
            ("", "dquote"),
            ("", "dquote")]]]),
    ("%", [
        ("", "dquote"),
        ("", "dquote"),
        ("", "dquote")]),
    ("%", [
        ("*", "ws"),
        ("", "eol")]]]),
("quotepar", [
    ("", [
        ("?", "r"),
        ("", "quote"),
        ("", "quote"),
        ("", "quote")]),
    ("*", [
        ("^", [
            ("", "quote"),
            ("", "quote"),
            ("", "quote")]]]),
    ("", [
        ("", "quote"),
        ("", "quote"),
        ("", "quote")]]]),
("quotepar", [
    ("", [
        ("?", "r"),
        ("", "dquote"),
        ("", "dquote"),
        ("", "dquote")]),
    ("*", [
        ("^", [
            ("", "dquote"),
            ("", "dquote"),
            ("", "dquote")]]]),
    ("", [

```

```
        ("", "dquote"), 208
        ("", "dquote"), 209
        ("", "dquote"])]), 210
("quotepar", [ 211
    ("", [ 212
        ("?", "r"), 213
        ("", "bquote"), 214
        ("", "bquote"), 215
        ("", "bquote"]]), 216
    ("*", [ 217
        ("^", [ 218
            ("", "bquote"), 219
            ("", "bquote"), 220
            ("", "bquote"])]]), 221
    ("", [ 222
        ("", "bquote"), 223
        ("", "bquote"), 224
        ("", "bquote"])]]), 225
("quotestring", [ 226
    ("?", "r"), 227
    ("", "quote"), 228
    ("*", [ 229
        ("^", "quote"]]), 230
    ("", "quote"]]), 231
("quotestring", [ 232
    ("?", "r"), 233
    ("", "dquote"), 234
    ("*", [ 235
        ("^", "dquote"]]), 236
    ("", "dquote"]]), 237
("quotestring", [ 238
    ("?", "r"), 239
    ("", "bquote"), 240
    ("*", [ 241
        ("^", "bquote"]]), 242
    ("", "bquote"]]), 243
("bol", [ 244
    ("", "bol"]]), 245
("eol", [ 246
    ("", "eol"]]), 247
("keyword", [ 248
    ("", "kw"]]), 249
("comment", [ 250
    ("", "comchar"), 251
    ("*", [ 252
        ("^", "eol"])]]), 253
("function", [ 254
    ("", "def"), 255
    ("", "ws"), 256
    ("", "name"]]), 257
("class", [ 258
    ("", "cls"), 259
    ("", "ws"), 260
    ("", "name"]]), 261
("importstm", [ 262
    ("", "import"), 263
    ("", "ws"), 264
    ("", [ 265
```



```

    "~": "{\\aatilde}",
    "/": "{\\aabar}",
    "<": "{\\aalt}",
    ">": "{\\aagt}",
    "*": "{\\aast}"

```

""" The mapping from source tokens to document dependent tokens. The substrings (except token) are fetched from the option dictionary in this class. The first element in the tuple is the new name of the token (after the map). """

```

tokmap = {
    "bod": ("", ("\\documentclass%(clsopt)s%(cls)s\\n%(preamble)s" +
                "\\begin{document}\\n\\title%(title)s\\n%(token)s")),
    "eod": ("", "%(token)s\\end{document}\\n"),
    "bof": ("", "\\begin{aasrc}\\n%(token)s"),
    "eof": ("", "%(token)s\\end{aasrc}\\n"),
    "bol": ("bol", "%(token)s"),
    "eol": ("eol", "%(token)s"),
    "ws": ("ws", "\\aaws%(token)s")}

```

""" The mapping from language to document components. Be aware that there should be a match between the number of substrings (the document format presented here and the the number of tuppels in the description list for the coresponding source language component (Src.srccomp). """

```

compmap = {
    "docstring": "\\adoc{%%}{%%}{%%}%%\\endaadoc{%%}%%",
    "quotepar": "\\aaqt{%%%%}",
    "quotestring": "\\aaqt{%%%%%%}",
    "bol": "\\aaline{%%}",
    "eol": "}%s",
    "keyword": "\\aakw{%%}",
    "comment": "%s\\aacom{%%}",
    "function": "\\aakw{%%}%%\\aafunc{%%}",
    "class": "\\aakw{%%}%%\\aacls{%%}",
    "importstm": "\\aakw{%%}%%\\aamod{%%}",
    "fromstm": "\\aakw{%%}%%\\aamod{%%}%%\\aakw{%%}%%\\aanm{%%}"}

```

class SrcTok:

""" Generate tokens from the source. """

```
def __init__(self, input=None, src=None, doc=None):
```

""" Initialize the SrcTok class. """

```
if input:
```

```
    self.input = input
```

```
else:
```

```
    self.input = sys.stdin
```

```
if src:
```

```
    self.src = src
```

```
else:
```

```
    self.src = Src()
```

```
if doc:
```

```
    self.doc = doc
```

```
else:
```

```
    self.doc = Doc()
```

```
self.compiledtokens = self.compileSrcTokens(self.src.srctok)
```

```
self.compiledall = regex.compile(string.join(map(
```

```
    lambda tokenit: tokenit[1][0],
```

```
    self.src.srctok), "\\|"))
```

```
self.linenum = 1
```

```
while self.linenum < string.atoi(self.src.option["startline"]):
```

```
    if not self.input.readline(): break
```

```

        self.linenum = self.linenum + 1                                404

def compileSrcTokens(self, srctok):                                  405
    """ Make a compiled srctokens tree. """                          406
    compiled = []                                                  407
    for (name, spec) in srctok:                                     410
        compiled.append((name, regex.compile(spec[0]),             411
                        self.compileSrcTokens(spec[1])))           412
    return compiled                                               413
                                                                    414
def searchToken(self, line, pos):                                  415
    """ Search for next token. """                                  416
    return self.compiledall.search(line, pos)                       417
                                                                    420
def bestToken(self, line, pos, compiled=[]):                      421
    """ Find the best token match (most specialized) at this position. """ 422
    if not compiled: compiled = self.compiledtokens                423
    for token in compiled:                                         426
        length = token[1].match(line, pos)                        427
        if length > 0:                                            428
            if token[2]:                                          429
                (nlength, nname) = self.bestToken(line, pos, token[2]) 430
                if nlength == length: return (nlength, nname)      431
            return (length, token[0])                              432
    return (0, "")                                                433
                                                                    434
def decompLine(self, tokens, line):                                435
    """ Decompose a line to a list of tokens. """                  436
    tabsize = string.atoi(self.src.option["tabsize"])              437
    linepos = index = 0                                           440
    pos = self.searchToken(line, index)                            441
    while pos != -1:                                               442
        (length, name) = self.bestToken(line, pos)                443
        if pos > index:                                           444
            tokens.append(("", line[index:pos]))                   445
        if name == "ws":                                          446
            num = 0                                               447
            for wsi in range(pos, pos + length):                   448
                if line[wsi] == "\t":                             449
                    num = num + (tabsize - ((linepos + num) % tabsize)) 450
                else:                                             451
                    num = num + 1                                  452
            linepos = linepos + num                                453
            token = 'num'                                          454
        else:                                                      455
            linepos = linepos + length                              456
            token = line[pos:pos+length]                           457
        tokens.append((name, token))                               458
        index = pos + length                                       459
        pos = self.searchToken(line, index)                        460
    if index < len(line):                                         461
        tokens.append(("", line[index:len(line)]))                462
                                                                    463
def fetchTokens(self, tokens):                                    464
    """ Fetch a new line from the source and decompose it to tokens. Returns false if there ain't no line 465
    to fetch. """                                                 466
    if (string.atoi(self.src.option["stopline"]) == -1 or         470
        self.linenum <= string.atoi(self.src.option["stopline"])): 471
        line = self.input.readline()                               472

```

```

        debug.write("%s." % 'self.linenum')
        self.linenum = self.linenum + 1
    else:
        line = ""
    if not line: return 0
    tokens.append(("bol", 'self.linenum - 1'))
    self.decompLine(tokens, line[:-1])
    tokens.append(("eol", "\n"))
    return 1

class Tok:
    """ A class to manage tokens (with possible check points). """

    def __init__(self, srctok=None, pre=[], post=[]):
        """ Initialize the Tok class. """
        if srctok:
            self.srctok = srctok
        else:
            self.srctok = SrcTok()
        self.src = self.srctok.src
        self.doc = self.srctok.doc
        self.tokenpos = -1
        self.nextlist = []
        if pre:
            self.tokens = pre
        else:
            if self.doc.option["document"] != "0":
                self.tokens = [("bod", "")]
                debug.write("Generating document\n")
            else:
                debug.write("Generating environment\n")
                self.tokens = [("bof", "")]
        if post:
            self.posttok = post
        else:
            if self.doc.option["document"] != "0":
                self.posttok = [("eod", "")]
            else:
                self.posttok = [("eof", "")]

    def checkPoint(self):
        """ Make a check point. We must save the position off current token so we can rollback. """
        self.nextlist.append(self.tokenpos)

    def commit(self):
        """ Ok, we committed the sequence of tokens from last check point. """
        del self.nextlist[-1]

    def rollBack(self):
        """ Don't commit the token sequence. Rollback to the last check point. """
        self.tokenpos = self.nextlist[-1]
        del self.nextlist[-1]

    def next(self):
        """ Fetch the next token. We may have to fetch a new line from the source file (with fetchTokens) """
        if self.tokenpos + 1 < len(self.tokens):
            self.tokenpos = self.tokenpos + 1

```



```
    else:
        if self.nextlist:
            self.tokenpos = self.tokenpos + 1
        else:
            self.tokens = []
            self.tokenpos = 0
        if not self.srctok.fetchTokens(self.tokens):
            if self.posttok:
                self.tokens = self.tokens + self.posttok
                self.posttok = []
            else:
                raise IndexError

    def current(self):
        """ Returns the current token. """
        return self.tokens[self.tokenpos]

class TokComp:
    """ Find components. """

    def __init__(self, tok=None):
        """ Initialize the TokComp class. """
        if tok:
            self.tok = tok
        else:
            self.tok = Tok()
        self.src = self.tok.src
        self.doc = self.tok.doc

    def mapChars(self, text, start=0, stop=0):
        """ Using charmap to map illegal document cahracters to commands. """
        if stop == 0: stop = len(text)
        ttext = ""
        for index in range(start, stop):
            try:
                ttext = ttext + self.doc.charmap[text[index]]
            except KeyError:
                ttext = ttext + text[index]
        return ttext

    def mapTok(self, mod):
        """ Using tokmap to map language tokens to document tokens. """
        (name, text) = self.tok.current()
        currcomp = self.src.option["currcomp"]
        if name == "bol" and not currcomp == "bol":
            try:
                tt = string.split(self.doc.compmmap[currcomp], "%s")[0]
            except KeyError:
                tt = ""
            ttext = self.doc.compmmap[name] % (text,) + tt
            if "%" in mod or "/" in mod:
                ttext = "" # Discard line number
            elif "#" in mod:
                ttext = text
            return ttext
        if name == "eol" and not currcomp == "eol":
            try:
                tt = string.split(self.doc.compmmap[currcomp], "%s")[-1]
```

```

        except KeyError:
            tt = ""
            ttext = tt + self.doc.compmap[name] % (text,)
            if "%" in mod or "/" in mod:
                ttext = text
            return ttext
    if "#" in mod or "%" in mod:
        if name == "ws":
            text = " " * string.atoi(text)
        return text
    if name == "ws" and "/" in mod:
        text = " " * string.atoi(text)
        return text
    try:
        self.doc.option["token"] = self.mapChars(text)
        return self.doc.tokmap[name][1] % self.doc.option
    except KeyError:
        return self.mapChars(text)

def matchComp(self, comp):
    """ Find one component matching tokens. The modifier specifies the type of the match (not, zero
    or more, single). Be aware that some modifiers are only interpreted by matcCompList. """
    self.tok.checkPoint()
    self.tok.next()
    text = ""
    try:
        if "^" in comp[0]:
            if self.tok.current()[0] != comp[1]:
                text = self.mapTok(comp[0])
            else:
                raise NoSymbolMatch
        elif "*" in comp[0]:
            self.tok.rollback() # We can match * with zero tokens
            self.tok.checkPoint()
            try:
                while 1:
                    self.tok.checkPoint()
                    self.tok.next()
                    if self.tok.current()[0] != comp[1]:
                        self.tok.rollback()
                        break
                    else:
                        self.tok.commit()
                        text = text + self.mapTok(comp[0])
            except IndexError:
                pass
        else:
            if self.tok.current()[0] == comp[1]:
                text = self.mapTok(comp[0])
            else:
                raise NoSymbolMatch
    except NoSymbolMatch:
        self.tok.rollback()
        raise NoSymbolMatch
    else:
        self.tok.commit()
        return text

```

```
def matchCompList(self, comp): 674
    """ Find a list component matching tokens. The modifier specifies the type of the match (not, zero
    or more, single, ...). """
    self.tok.checkPoint() 679
    text = "" 680
    try: 681
        # ? and + are the same for both lists and strings 682
        if "?" in comp[0]: 683
            try: 684
                text = self.matchCompList((comp[0][1:], comp[1])) 685
            except NoSymbolMatch: 686
                pass 687
        elif "+" in comp[0]: 688
            try: 689
                text = self.matchCompList((comp[0][1:], comp[1])) 690
            except NoSymbolMatch: 691
                raise NoSymbolMatch 692
        else: 693
            try: 694
                ttext = self.matchCompList( 695
                    ("*" + comp[0][1:], comp[1])) 696
            except NoSymbolMatch: 697
                pass 698
            else: 699
                text = text + ttext 700
        # Use matchComp if the component is not a list 701
        elif type(comp[1]) is StringType: 702
            try: 703
                text = self.matchComp(comp) 704
            except NoSymbolMatch: 705
                raise NoSymbolMatch 706
        # Ok, we know it is a list 707
        elif "/" in comp[0]: 708
            for (mod, ccomp) in comp[1]: 709
                try: 710
                    text = self.matchCompList((mod + comp[0][1:], ccomp)) 711
                except NoSymbolMatch: 712
                    continue 713
                else: 714
                    break 715
            else: 716
                raise NoSymbolMatch 717
        elif "^" in comp[0]: 718
            self.tok.checkPoint() 719
            for (mod, ccomp) in comp[1]: 720
                try: 721
                    ttext = self.matchCompList((mod + comp[0][1:], ccomp)) 722
                except NoSymbolMatch: 723
                    self.tok.next() 724
                    text = text + self.mapTok(comp[0]) 725
                    self.tok.commit() 726
                break 727
            else: 728
                text = text + ttext 729
        else: 730
            self.tok.rollback() 731
            raise NoSymbolMatch 732
    elif "*" in comp[0]: 733
```

```

        cont = 1
        while cont:
            cont = 0; ttext = ""
            self.tok.checkPoint()
            for (mod, ccomp) in comp[1]:
                try:
                    ttext = self.matchCompList(
                        (mod + comp[0][1:], ccomp))
                except NoSymbolMatch:
                    self.tok.rollback()
                    break
                else:
                    ttext = ttext + ttext
            else:
                self.tok.commit()
                text = text + ttext
                cont = 1
        else:
            if comp[0]:
                mmod = comp[0]
            else:
                mmod = ""
            for (mod, ccomp) in comp[1]:
                try:
                    ttext = self.matchCompList((mod + mmod, ccomp))
                except NoSymbolMatch:
                    raise NoSymbolMatch
                else:
                    text = text + ttext
        except NoSymbolMatch:
            self.tok.rollback()
            raise NoSymbolMatch
        else:
            self.tok.commit()
            return text

class Comp:
    """ Generates components (a name and a texttuple) from tokens. """

    def __init__(self, tokcomp=None):
        """ Initialize the TokComp class. """
        if tokcomp:
            self.tokcomp = tokcomp
        else:
            self.tokcomp = TokComp()
        self.tok = self.tokcomp.tok
        self.src = self.tokcomp.src
        self.doc = self.tokcomp.doc
        self.ftokcomp = {}
        for index in range(len(self.src.srccomp)):
            tlist = self.src.srccomp[index][1]
            try:
                self.appendFtok(index, tlist)
            except KeyError:
                self.ftokcomp = {}
                break
        if self.ftokcomp:
            debug.write("Using speedup\n")

```

```

796
def appendFtok(self, index, tlist):
797
798     if tlist[0][0] != "":
799         raise KeyError
800
801     if type(tlist[0][1]) is StringType:
802         try:
803             self.ftokcomp[tlist[0][1]].append(index)
804         except KeyError:
805             self.ftokcomp[tlist[0][1]] = [index]
806
807     else:
808         self.appendFtok(index, tlist[0][1])
809
810 def fetchComp(self, comp):
811     """ fetch a component by name. """
812     self.tok.checkPoint()
813     texttuple = ()
814     try:
815         for (mod, ccomp) in comp:
816             try:
817                 text = self.tokcomp.matchCompList((mod, ccomp))
818             except NoSymbolMatch:
819                 self.tok.rollback()
820                 raise NoSymbolMatch
821             else:
822                 texttuple = texttuple + (text,)
823         else:
824             self.tok.commit()
825             return texttuple
826     except (KeyError, NoSymbolMatch):
827         raise NoSymbolMatch
828
829 def searchComp(self):
830     """ Search for components from in src.srccomp. Use speedup if available (based on value of first
831     token). """
832     comprange = []
833     if self.ftokcomp:
834         self.tok.checkPoint()
835         self.tok.next()
836         try:
837             comprange = self.ftokcomp[self.tok.current()[0]]
838         except KeyError:
839             pass
840         self.tok.rollback()
841     if not comprange:
842         comprange = range(len(self.src.srccomp))
843     for index in comprange:
844         (name, comp) = self.src.srccomp[index]
845         self.src.option["currcomp"] = name
846         try:
847             texttuple = self.fetchComp(comp)
848         except NoSymbolMatch:
849             continue
850         else:
851             return (name, texttuple)
852     self.tok.next()
853     self.src.option["currcomp"] = ""
854     return ("", (self.tokcomp.mapTok(""),))
855
856
857

```

```
class CompDoc: 858
    """ Map characters, tokens and components to document format. """ 859
    862
    def __init__(self, output=None, comp=None): 863
        """ Initialize the CompDoc class. """ 864
        if output: 867
            self.output = output 868
        else: 869
            self.output = sys.stdout 870
        if comp: 871
            self.comp = comp 872
        else: 873
            self.comp = Comp() 874
        self.doc = self.comp.doc 875
    876
    def mapComp(self, name, texttuple): 877
        """ Using compmap to map language components to document components. """ 878
        try: 881
            return self.doc.compmap[name] % texttuple 882
        except KeyError: 883
            return string.joinfields(texttuple, "") 884
    885
    886
    def printDoc(self): 887
        debug.write("Starting\n") 888
        while 1: 889
            try: 890
                (name, texttuple) = self.comp.searchComp() 891
                sys.stdout.write(self.mapComp(name, texttuple)) 892
            except IndexError: 893
                break 894
        debug.write("\nDone\n") 895
    896
class Debug: 897
    """ Handy stuffs """ 898
    901
    def __init__(self): 902
        """ Initialize the Debug class. """ 903
        self.write = sys.stderr.write 906
    907
# Make an instance of Debug 908
debug = Debug() 909
    910
    911
""" If this is the main file, we do it with default values. """ 912
if __name__ == "__main__": 915
    CompDoc().printDoc() 916
```